

大数据提升灾难信息管理水平

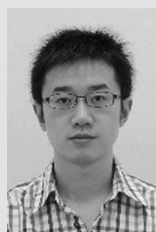
Big Data Meets the Needs of Disaster Information Management



李涛, 男, 南京邮电大学计算机学院、软件学院院长, 南京邮电大学大数据研究院院长。2004年7月获美国罗彻斯特大学 (University of Rochester) 计算机科学博士学位, 2004-2014年先后任美国佛罗里达国际大学 (Florida International University) 计算机学院助理教授、副教授 (终身教授)、教授 (full professor)、研究生主管 (graduate program director)。由于在数据挖掘及应用领域成效显著的研究工作, 曾多次获得各种荣誉和奖励, 其中包括2006年美国国家自然科学基金委员会颁发的杰出青年教授奖, 2010年IBM大规模数据分析创新奖, 并于2009年获得佛罗里达国际大学最高学术研究奖。



周绮凤, 女, 博士, 厦门大学自动化系副教授。2002年起从事数据挖掘及智能系统方面的研究工作, 2014-2015年在美国佛罗里达国际大学访学, 主要研究兴趣包括机器学习、数据挖掘及其在可持续发展等领域的应用。



郑理, 男, 2014年在美国佛罗里达国际大学获得计算机科学博士学位, 目前是南京邮电大学计算机学院项目研究员。主要研究兴趣包括信息检索、推荐系统及灾难信息管理, 发表多篇顶级数据挖掘国际期刊和会议论文, 参与多本数据挖掘相关应用领域书籍编写。



黄越, 男, 南京邮电大学科技处助理研究员, 主要研究方向为大数据、数据挖掘。

doi: 10.11959/j.issn.2096-0271.2015035

1 灾难信息管理

灾难管理旨在有效地应对和避免自然灾害(如飓风、地震、海啸、火灾)及人为灾害(如战争、恐怖袭击)等紧急事件给社会和民众带来的财产损失和生命威胁^[1]。近年来,随着自然灾害的不断发生、人为破坏和恐怖主义的蔓延,灾难管理和灾难恢复受到了越来越多的关注。如何能够快速准确地预测灾难发生的方式和类型,评估灾难的破坏程度和影响以及制定灾后恢复的方针和措施,对保护国家和公民的财产和生命安全、减少灾难影响和损失、提高灾后重建的资源利用和整合效率,都起到了至关重要的作用。

灾难管理作为一个庞大的管理体系与整个国民经济息息相关。政府间各个部门、非政府组织、民间团体甚至个人需要紧密合作,建立顺畅的信息沟通渠道和合理的资源共享平台。近年来,信息的爆炸式增长使得原来单纯依靠收集、存储和查询数据的简单管理方式在大数据环境下(数据量大、形式复杂、实时性强)变得不再实用。因此,迫切需要有效的数据处理和分析手段将有价值的信息从不断增长的海量数据中提取出来。大数据挖掘技术建立了从数据到信息再到知识的转化流程,提供了高效处理复杂数据的工具和方法,因而具备很强的应用潜力,能够将灾难信息管理水平提升到一个新的台阶。比如,谷歌公司(Google)通过从全球的博客(Blog)中挖掘出和流感相关的信息,从而建立了一个预警机制¹。可以预见,越来越多的先进数据分析技术将被运用到该领域中,从而推动灾难管理水平的提升。

2 大数据提升灾难信息管理

基于大数据的灾难管理已成为新一代灾难管理和应急处理领域的核心前沿研究课题^[2]。一方面,基于大数据的灾难信息系统研究,在国际社会尤其是欧美发达国家得到了极大的重视和推动,如2014年,美国政府和日本政府联合发布了利用大数据技术来帮助灾难研究的专项研究(US-Japan Big Data and Disaster Research (BDD)²)计划。另一方面,灾难管理领域中信息的独特性使该领域中的信息管理、处理和分析面临很大的挑战。通过已有的研究,将灾难管理领域中数据处理的难点总结为以下6个方面^[1]:数据爆炸但知识匮乏、信息冗余、信息不一致、时间和地理位置敏感、用户角色复杂和领域知识的使用。一套成功的灾难信息管理系统应该能够有针对性地应对和解决以上信息传递中的数据特点和难点,最大程度地保证灾难信息管理和共享平台的及时性、有效性和可靠性,从而达到灾难管理的最终目标,即在正确的时间给正确的人传递正确的信息。

3 一个基于大数据技术的灾难信息管理成功项目案例

笔者领导的佛罗里达国际大学的科研团队与南佛罗里达地区的灾难应急管理中心通过建立长期合作关系,共同研究在灾难信息管理领域内的数据收集、整理、存储和分析技术的应用^[3~10]。该项目中参与的机构包括:迈阿密戴德县(Miami-Dade County)灾难应急管理中心(Emergency Operation Center, EOC)、Wal-mart、

1
<http://www.google.org/flutrends>

2
<http://www.nsf.gov/pubs/2014/nsf14575/nsf14575.htm>

Home Depot、Verizon、Ryder等。目前，通过与美国联邦紧急事务管理署（Federal Emergency Management Agency, FEMA）等多方合作，基于Web的系统原型——商业连续与灾难恢复信息管理系统，已经上线（www.bizrecovery.org）并在迈阿密和塔拉哈西建立镜像。

这个系统主要用于灾难恢复阶段的社区管理、信息分类、信息共享以及灾难数据的收集。图1是原型系统的截图。图1上半部分是信息板，主要用于展示最近发生的重要事件和用户订阅的特定类型的事件信息，如机场、学校和交通的当前状态。每行代表一个在地理位置上相对独立的行政区域，每列代表一个事件的实体类别，比如

第一列的第二个圆形深色标记提示最新发生的棕榈滩机场关闭信息。提示板功能的实现是通过提取文本消息中的地理位置信息、时间信息、事件实体并关联相应的状态信息，最终转换为结构化的数据进行存储和展示。左下部分提供当前灾难事件的全面展示，如图片、视频、文档等；右下部分展示了系统收集和用户提交的相关消息，支持时间、地点等多种方式的信息过滤。

在用户使用的终端设备方面，传统的固定终端（如台式个人电脑）在获取信息和上传数据的实时性方面受到限制，尤其是在灾难发生时。因此，研究小组设计开发了一款基于iOS移动平台的灾难信息管

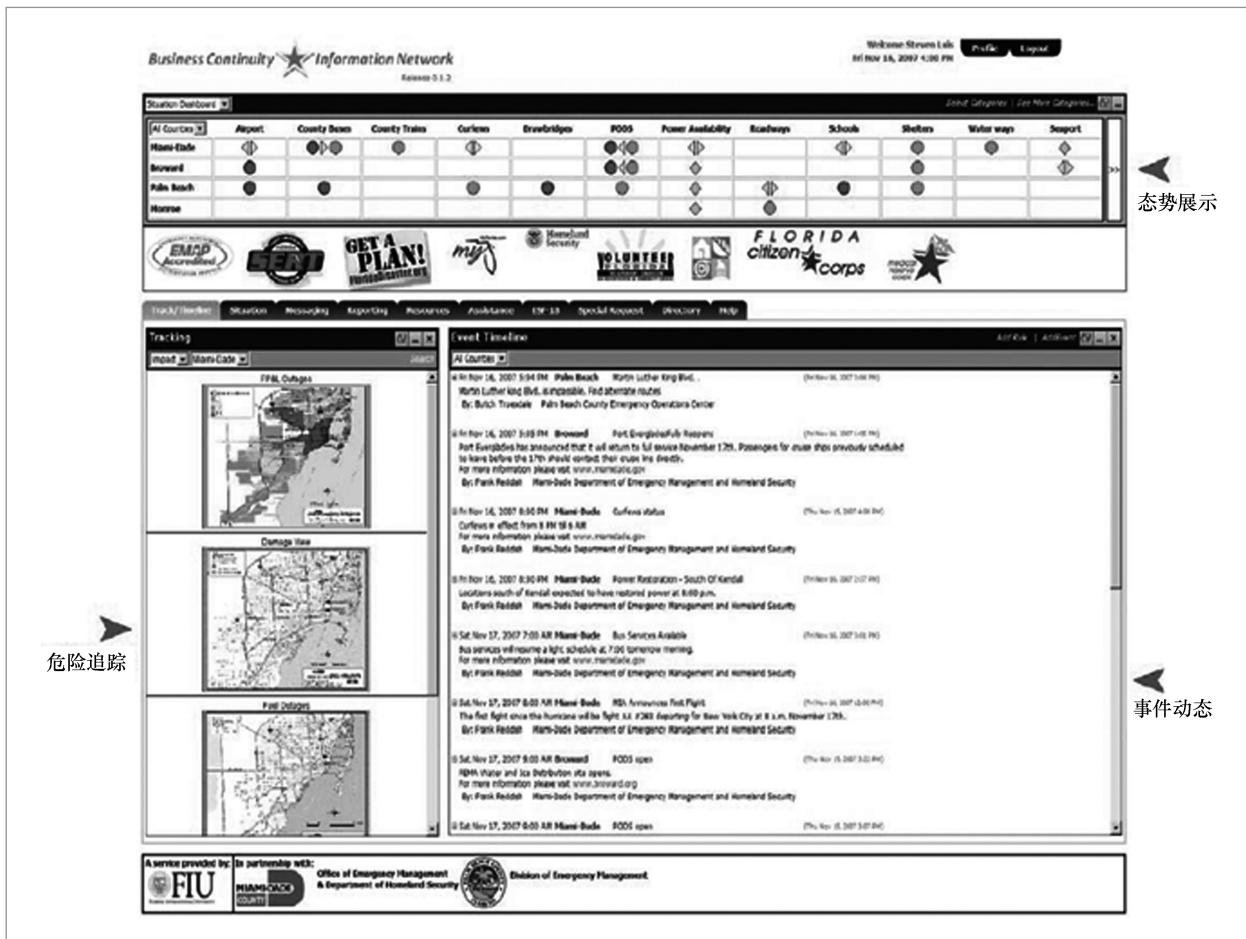


图1 Web 系统界面

理应用——ADSB (all-hazard disaster situation browser, 全息灾害灾情浏览器), 系统架构如图2所示^[11]。它可以运行在手持移动设备上, 支持灾难事件消息的阅读、上传、推荐、文摘和个性化社区管理等多种功能, 如图3所示。

这些系统平台和应用利用了非常广泛的数据资源, 涵盖了静态数据和动态数据两大类。静态数据资源包括: 当地历史灾难管理数据; 在灾难准备、灾难应急和灾难恢复阶段的应急指南和行动手册; 企业和重要设施的地理位置; 公共的地理信息数据以及交通网络。动态数据资源包括: 合作机构在灾难事件期间的报告 (包括当前的灾难威胁状况、灾难准备进展和总体的应对灾难的目标和策略); 损失分析评估报告和图片; 关键交通枢纽 (公路、高速公路、桥梁、港口等) 的状态; 重要基础设施 (能源、电力、运输等) 的状态; 应急服务 (消防、治安、医疗等) 的状态和重要公共设施 (学校等) 的状态; 不同媒体的新闻报道; 电子邮件、邮件列表、发布会或会议内容等; 灾难呼叫中心关于损失状态报告的接入拨打日志; 社交网站、Blog、Twitter 的数据信息。除以上数据来源之外, 还筛选出一组与灾难信息高度相关的网站链接作为可靠信息源, 借助网络爬虫来获取互联网数据。

在灾难管理领域的研究和实践中, 总结出了如下5个应用大数据挖掘技术来保证灾难管理系统平台成功发挥功效的关键要素: 迅速获得相关资源; 准确提取相关信息; 合理组织相关信息; 有效管理用户角色和资料; 及时发现和组织社区。通过针对性地把控和解决上述关键要素, 最终实现了灾难管理系统的“有效获取灾难事件的相关信息, 提高在复杂信息环境下的觉察力”、“自动获取用户关注点并有效地传递相关信息”以及“更好地利用社区信

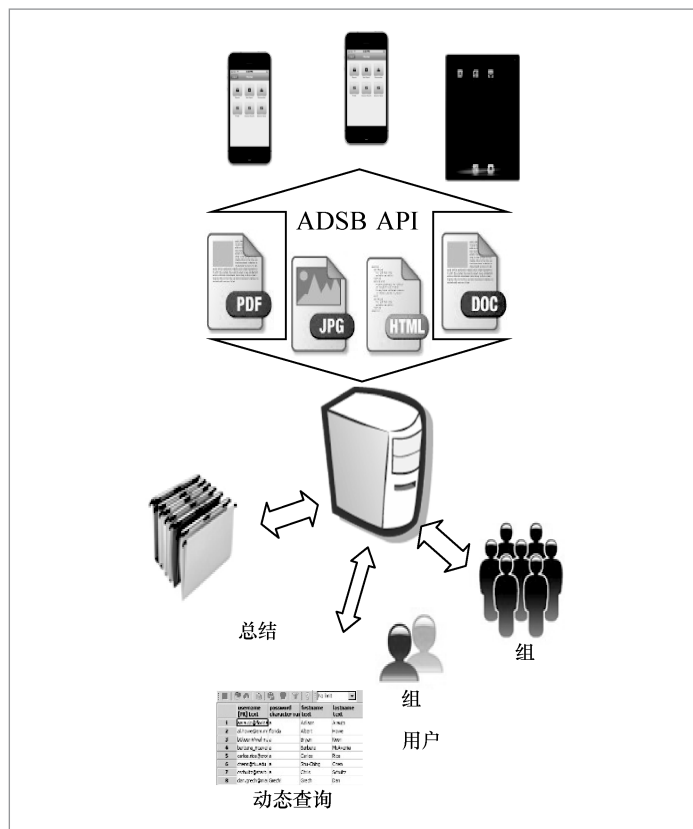


图2 ADSB 系统架构

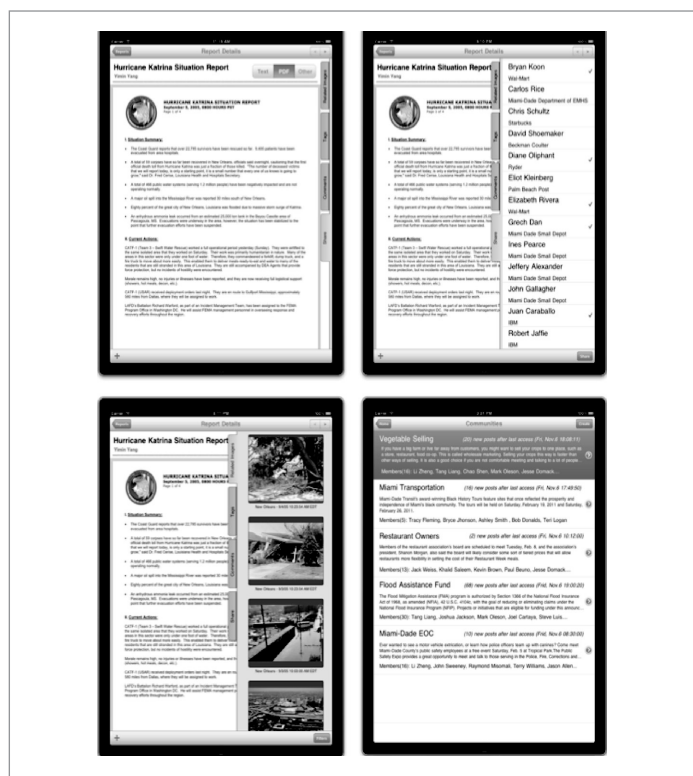


图3 ADSB 系统界面

息进行灾难恢复”这三大目标任务,从而有效地支持公共机构和私有部门之间更好地进行信息交互和共享。

大数据时代的灾难信息管理已经成为一个理论研究和实际应用紧密结合、社会政府及个人都迫切需求的重要研究内容。其中,笔者带领的研究团队关于大数据挖掘在灾难信息管理方面的研究受到国内外学术界的广泛关注以及政府部门的高度重视。开发的系统是一个利用大数据技术,包括信息抽取、信息检索、信息过滤和决策支持等,有效提升灾难信息获取、灾难应急管理等方面能力的成功案例,也是第一个基于Web的政府与企业进行沟通、交流,实现信息共享、事物协作等的工具。目前已有American Airlines、Aon Corporation、AT&T、Bank of America、IBM等60多家企业和部门加入了这个信息网络。由于该项目的独创性及成功应用,开发的系统被美国联邦紧急事务管理署列为政府与企业成功合作的一个典范³。

3

http://users.cis.fiu.edu/~chens/docs/miami_dwindow.onresize=myChart1.resize;ade_partnership.pdf

4 结束语

与国外相比,国内目前在灾难信息管理及相关领域的研究相对较少,尚缺乏全面、综合、有效的灾难信息处理系统和典型示范应用。因此,如何将基于大数据的灾难信息管理在国内进行推广,积极开展基于大数据的智能化灾害管理与应急处理关键技术及系统研制,具有重要的研究意义和重大的社会价值。值得注意的是,该项推广不是上述案例的简单复制和扩展,而是需要结合国内灾难信息管理的特点以及不同地区灾难管理的任务目标、信息需求,开展有特色的、个性化的大数据灾难信息管理开发和应用。

参考文献

- [1] 李涛. 数据挖掘的应用与实践. 厦门: 厦门大学出版社, 2013
Li T. Data Mining Where Theory Meets Practice. Xiamen: Xiamen University Press, 2013
- [2] Hristidis V, Chen S C, Li T, *et al.* Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 2010, 83(10):1701~1714
- [3] Zheng L, Shen C, Tang L, *et al.* Data mining meets the needs of disaster information management. *IEEE Transactions on Human-Machine Systems*, 2013, 43(5): 451~464
- [4] Li L, Li T. An empirical study of ontology-based multi-document summarization in disaster management. *IEEE Transactions SMC: Systems*, 2014, 44(2): 162~171
- [5] Hristidis V, Chen S C, Li T, *et al.* Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 2010, 83(10): 1701~1714
- [6] Zhou W B, Shen C, Li T, *et al.* Generating textual storyline to improve situation awareness in disaster management. *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration (IRI)*, Redwood City, CA, USA, 2014: 585~592
- [7] Zhou W B, Shen C, Li T, *et al.* A bipartite-graph based approach for disaster susceptibility comparisons among cities. *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration (IRI)*, Redwood City, CA, USA, 2014: 593~599
- [8] Zheng L, Shen C, Tang L, *et al.* Disaster sitrep—a vertical search engine and information analysis tool in disaster management domain. *Proceedings of the 13th IEEE International Conference on*

- Information Integration and Reuse (IRI), Las Vegas, USA, 2012: 457~465
- [9] Zheng L, Shen C, Tang L, *et al.* Using data mining techniques to address critical information exchange needs in disaster affected public-private networks. Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 2010: 125~134
- [10] Li L, Wang D D, Shen C, *et al.* Ontology-enriched multi-document summarization in disaster management. Proceedings of the 33rd International ACM SIGIR conference on Research and Development in Information Retrieval, New York, USA, 2010: 819~820
- [11] Zheng L, Shen C, Tang L, *et al.* Applying data mining techniques to address disaster information management challenges on mobile devices. Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, California, USA, 2011: 283~291 □